APPLICATION OF PARAMETRIC TIME SERIES ANALYSIS METHODS TO DESCRIPTIONS OF WAVE-INDUCED FLUCTUATIONS IN THE ADJACENT AIRFLOW

Craig Howard Smith

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

Application of Parametric Time Series Analysis Methods to Descriptions of Wave-Induced Fluctuations in the Adjacent Airflow

by

Craig Howard Smith

Thesis Advisor: K. L. Davidson

September 1972

Approved for public release; distribution unlimited.

Application of Parametric Time Series Analysis
Methods to Descriptions of Wave-
Induced Fluctuations in the Adjacent Airflow

by

Craig Howard Smith
Ensign, United States Navy
A.B., Princeton University, 1971

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

NAVAL POSTGRADUATE SCHOOL
September 1972

ABSTRACT

Certain methods of analysis for testing the hypothesis that organized motion exists in the airflow above naturally occurring water waves are examined. In particular, two current methods, spectral analysis and joint probability density function analysis, are briefly discussed; two new methods, matched filters and parametric time series analysis, are suggested.

An analysis is done with parametric time series analysis on wind-wave data. The results show the tractability of this data to parametric methods, and the results are in agreement with previous spectral analyses of the data in the regions of overlap.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$a(t)$	discrete time series of independent random shocks
a_t	an element of $a(t)$, the observation at time t
AR(p)	autoregressive model of order p
ARIMA(p,d,q)	autoregressive integrated moving average model of order (p,d,q)
ARIMA(p,d,q)x(P,D,Q)	multiplicative autoregressive integrated moving average of order (p,d,q)x(P,D,Q)
B	backward shift operator, such that $BX_t = X_{t-1}$
b	adaptive coefficient in Fourier series
d	integer exponent of ∇^d operator
D	as operator, $D \equiv d/dt$ as exponent, integer exponent of ∇_s^D operator
H	unknown coefficient in transfer function model
i	positive integer
M	integer parameter; M=0 if z(t) series is used in analysis, M=1 if $\tilde{z}(t)$ series is used
MA(q)	moving average model of order q
N	number of observations in z(t) less d
N_t	structured noise component in transfer function model
p	order parameter of AR(p) model
P	order parameter of seasonal AR component of ARIMA(p,d,q)x(P,D,Q)model
q	order parameter of MA(q) model

Q	order parameter of seasonal MA component of ARIMA $(p,d,q)x(P,D,Q)$ model
$r_m(a)$	autocorrelation of $a(t)$ series at lag m
S	inverse of ∇ operator
s	integer number of lags per period of a seasonal effect
$v(B)$	equivalent filter series for transfer function models
$w(t)$	time series obtained by application of $\nabla^d \nabla_s^D$ operator to $z(t)$
w_t	an element of $w(t)$, the observation at time t
$X(t)$	an observed time series
X_t	an element of $X(t)$: observation at time t
$Y(t)$	an observed time series
Y_t	an element of $Y(t)$: observation at time t
$z(t)$	an observed discrete time series
z_t	element of $z(t)$: observation at time t
$\tilde{z}_t$	$\tilde{z}_t = z_t - \mu$
δ	unknown coefficient in transfer function
$\delta(B)$	operator with unknown coefficients in transfer function model
η	unknown coefficient in transfer function model
θ_i	i^{th} unknown coefficient of $\theta(B)$
Θ_i	i^{th} unknown coefficient of $\Theta(B)$
$\theta(B)$	moving average operator
$\Theta(B^s)$	seasonal moving average operator

μ	mean of z(t) series
ξ	unknown parameter in transfer function models
Ξ	unknown parameter in transfer function models
σ_a^2	variance of a(t) series
Σ	summation operator
τ	dead-time of a system
ϕ_i	i^{th} unknown parameter of $\phi(B)$
Φ_i	i^{th} unknown parameter of $\Phi(B)$
$\varphi(B)$	generalized autoregressive operator
$\phi(B)$	autoregressive operator
$\Phi(B^s)$	seasonal autoregressive operator
ψ_i	i^{th} unknown parameter in $\psi(B)$
$\psi(B)$	linear filter operator
ω_i	unknown coefficient in transfer function model
$\omega(B)$	operator with unknown coefficients in transfer function model
$\Omega(B)$	operator with unknown coefficients in transfer function model
∇^d	$\nabla^d = (1-B)^d$
∇_s^D	$\nabla_s^D = (1-B^s)^D$

ACKNOWLEDGMENT

The author extends his grateful appreciation to his advisor, Professor K. L. Davidson, for his unflagging patience and interest in this study, and his valuable help and advice in the preparation of this paper.

Special thank-you's are extended to my second reader, Professor P. C. C. Wang, for his useful advice and friendly counsel.

The author wishes to express here his appreciation of the help given him by Professor F. R. Richards, Mr. Jerry Learmouth, the W. R. Church Computer Facility employees, Professor G. A. Stoops, and Professor J. Stuart Hunter of Princeton University, without whose help this paper would have been a long time coming.

Finally, the author further extends a heartfelt thanks to his typist, Marion Marks. Without her outstanding efforts and willingness to work beyond the expected, this paper might even yet exist only as a handscrawled manuscript in the author's notebook.

I. INTRODUCTION

In recent years, interest in parameterizing and defining processes in the air-sea boundary layer has intensified. Recent theoretical models [Miles, 1957] postulate, among other things, that the airflow near the surface and the waves are coupled in such a manner as to produce organized motion in the air in relation to the wave field. This results in air motion with the same period as the dominant period of the waves. This paper examines certain methods of analysis by which these hypotheses can be tested. In particular, two current methods are discussed, two new methods suggested, and an analysis using one of the new methods is performed.

II. BACKGROUND

If organized motion exists in the air due to the presence of the underlying waves, its existence and nature should be reflected in observations of the wind over the waves. Unfortunately for a prospective analyst of such observations, this regime is characterized by the presence of a general turbulent field which serves to mask the organized motion. The problem, then, is to discover whether organized motion does exist, and if it does, its nature, in the midst of random turbulent influences. This should be determined from data obtained over natural waves.

It is the usual case that data of interest take the form of simultaneous continuous time traces of variables of interest, like wind components and wave heights. For purposes of ease of analysis, such records are usually subsequently discretized so that the final form is that of a multivariate discrete time series. This is the case with data used in this paper. That data was obtained from observations over natural waves on Lake Michigan; the data are fully described elsewhere [Davidson, 1970].

Time series differ from other types of statistical sample realizations in that the observations are assumed to be dependent upon one another, whereas in the larger body of statistical technique, it is preferred that observations be independent. The exploitation of this expected dependence is a distinguishing characteristic of time series analysis.

As is suggested by the form of the data, the majority of the techniques considered here are based on time series analysis. One method, that of joint probability distribution function analysis, is not drawn from the body of time series analysis, but it also uses the dependent nature of the data to advantage, though indirectly.

A. SPECTRAL ANALYSIS

Most widely known of the time series analysis techniques is spectral analysis. This technique examines the data in the frequency domain. The most common representation of spectral results is the sample spectrum, whose magnitude for a given frequency reflects the amount of variance in the data which can be accounted for by the presence of a cosine component at the given frequency. It is, in fact, the Fourier cosine transform of the estimated autocovariance function. Hence, the spectrum of data arising from a process dominated by a periodic effect should exhibit a marked peak at the frequency of the periodic effect, while the spectrum of white noise should be flat. Thus noise should raise the level of the spectrum, but should not obscure significant peaks. Generally, if the hypothesized relationship hold, one would expect a peak in the wind component spectra at the same frequency as the major peak in the wave spectrum.

Spectral analysis also has a natural extension into multivariate space, called cross-spectral analysis. As the covariance of wind and waves is of primary interest here, calculation of cross-spectra should reveal some information

of interest, including estimates of phase relationships, coherence, and cross-covariance with respect to frequency. Predictions from theoretical models for these values may then be compared to the estimates resulting from actual observations as an indication of the validity of the theory.

Spectral analysis is not without major drawbacks, however. A major handicap is that it can only deal adequately with stationary time series. Often natural time series are not observed to be stationary, and may require elaborate algorithms to be applied in an effort to remove trends and/or seasonality to obtain stationarity. A second difficulty is the lack of smoothness in actual sample spectra. A typical sample spectrum is so jagged that interpretation is difficult. The usual remedy is to smooth the data, smooth the spectrum, or both. This makes the information more readily apprehended, but the complex effect of such smoothing on individual estimates of spectral ordinates makes the statistical significance of the estimates difficult or impossible to determine. It may also result in a loss of information contained in the data. Further, while confidence intervals for unsmoothed estimates of spectral ordinates may be calculated theoretically, theory also shows that the error in the estimate of the spectral ordinate will be of the same order of magnitude as the spectral ordinate itself [Kendall and Stuart, 1966]. These facts make it difficult to quantitatively support the belief that a particular peak in a smoothed spectrum is significant. Thus Kendall and Stuart [1966] rightly caution

meteorologists and oceanographers by name about assuming the existence of periodic elements in the data generating process solely on the basis of spectral analysis; care is necessary. Further discussions of other possible pitfalls as well as much more complete discussions of spectral analysis are to be found in Blackman and Tukey [1958], Kendall and Stuart [1966], and Jenkins and Watts [1968].

Although unsatisfying in some statistical senses, spectral analysis is a useful tool to be applied to looking for wind-wave coupling. The data used in this paper was extensively analyzed spectrally by Davidson [1970], with good results.

B. MATCHED AND WIENER FILTERS

Another type of analysis related to spectral analysis which may hold promise for the wind-wave coupling problem is to be found in the repertoire of electrical engineering. Electrical engineers are frequently faced with the problem of detecting and isolating a signal in the midst of noise, precisely analogous to looking for organized motion in the air in the midst of turbulence. To solve this problem, two types of spectrally derived filters have been developed, designed to give a large response to signal and a small response to noise. They are matched filters and Wiener filters. To the best knowledge of the author, no one has yet applied such filters to wind-wave data looking for organized motion, but Sokol [1971] applied this technique to detection of thermal plumes in the air adjacent to the sea from temperature traces of a type similar to the wind-wave data. His results

were quite encouraging for the hypothesis that such an analysis might be useful in the detection of organized motion in turbulence.

At present, the major problem with the application of these techniques to wind-wave data appears to be sensitivity to small errors in estimates of certain parameters which are difficult to estimate accurately, like the phase of the periodic effect. However, the magnitude of such problems is as yet unknown, and does not preclude at least a comprehensive preliminary examination of these techniques for practicality with respect to the wind-wave problem.

C. JOINT PROBABILITY DENSITY FUNCTION ANALYSIS

The one analysis technique currently used in this field which is not properly a member of time series analysis is joint probability density function analysis. Basically, this method examines the data in the probability domain. In the analysis, the probability density function of the set of joint occurrences of two variables is calculated. Equal joint probability density contours are plotted on a two dimensional array. The contours are then subjectively analyzed for deviations from the concentric circle pattern of a theoretical random, independent joint probability density function, the current reference function being the bivariate standard normal joint probability density function. The pattern of these deviations, which is assumed to arise because of the dependent nature of the time series, is then

evaluated in terms of the generating process and compared to what is expected by theory. This type of analysis has been applied to meteorological data by Holland [1972], Davidson and Frank [1972], and Frank [1971].

This method shows great promise; however, it lacks refinement as yet. The choice of the bivariate standard normal joint probability density function as a reference distribution is formally incorrect, although it may be satisfactory in cases involving many degrees of freedom. The proper choice is a bivariate distribution analogous to the Student-t distribution in the univariate case. Using such a distribution, described in some detail in Birnbaum [1962], it appears quite feasible to develop an objective test of the significance of deviations from the distribution expected if the generating processes were random and independent. Further tests against such a reference distribution reduce, effectively, to t-tests for correlation coefficients and regression coefficients calculated between the two variables [Birnbaum, 1962, pages 223-242]. It would appear beneficial to include at least these last tests as a normal part of the analysis.

D. PARAMETRIC TIME SERIES ANALYSIS

The final method is that advanced by Box and Jenkins [1970], called by some, parametric time series analysis. This method involves fitting a stochastic model to the time series, while allowing, in keeping with modern statistical thought, the data themselves to shape the model and the analysis itself to a relatively great extent. This method is

relatively new; its capabilities and limitations are not common knowledge in the manner that those of spectral analysis are. In order to better understand these capabilities and limitations, particularly as they apply to the boundary layer data in this study, it is necessary to discuss the form of the models to be fitted. This is done in Sections 1 through 3, which describe, though with considerably less detail, the ideas expressed by the primary source, Box and Jenkins' text [1970]. With this background already given, the particular applications to wind-wave coupling investigations are discussed in Section E.

1. Yule's Proposition

Probably no process generating experimental observations is completely deterministic. Thus, deterministic models are not usually adequate in analyses, although physicists and others have successfully used them in analyses of data generated by an apparatus designed to be so accurate that noise contributions to the observations are negligible in relation to results of interest. But if the noise is of a significant magnitude, as with turbulence data, models including noise must be used. If the probability structure of the process is of interest, as in the case of atmospheric turbulence studies, the normal course is to use a stochastic model. The model fitted by parametric time series analysis is stochastic.

Central to parametric time series analysis is Yule's proposition [Yule, 1927] that a time series whose elements

are highly dependent, as is the case with wind-wave records, may usefully be viewed as having arisen from the application of a linear filter to a parallel series of independent random shocks drawn from a fixed distribution. This viewpoint is purely heuristic, and need not correspond to the actual process at all to be useful. The proposition is symbolized in Figure 1a.

In applying this idea, the fixed distribution of the series of random shocks will be taken to be normal with mean zero and variance σ_a^2 . With the random shock series represented by a_t's, the observed time series by z_t's, and the linear filter elements by ψ_i's, the relationship can be symbolized as

$$\begin{aligned} z_t &= \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \\ &= \mu + (1 + \psi_1 B + \psi_2 B^2 + \dots)\, a_t \\ &= \mu + \psi(B)\, a_t \end{aligned} \tag{2.1}$$

where B is the standard backward shift operator, and

$$B^s x_t = x_{t-s} \tag{2.2}$$

If the sequence $\psi(B)$ is finite, or infinite and convergent for B given a value of unity, the filter is said to be convergent. If the filter is convergent, then z(t) is a stationary time series, and μ is the mean of the series. If the filter is unstable, μ is an arbitrary reference point for the process level, and z(t) is non-stationary.

2. <u>The ARIMA Model</u>

Parametric time series analysis is based on fitting a particularly useful subset of linear filters of the kind

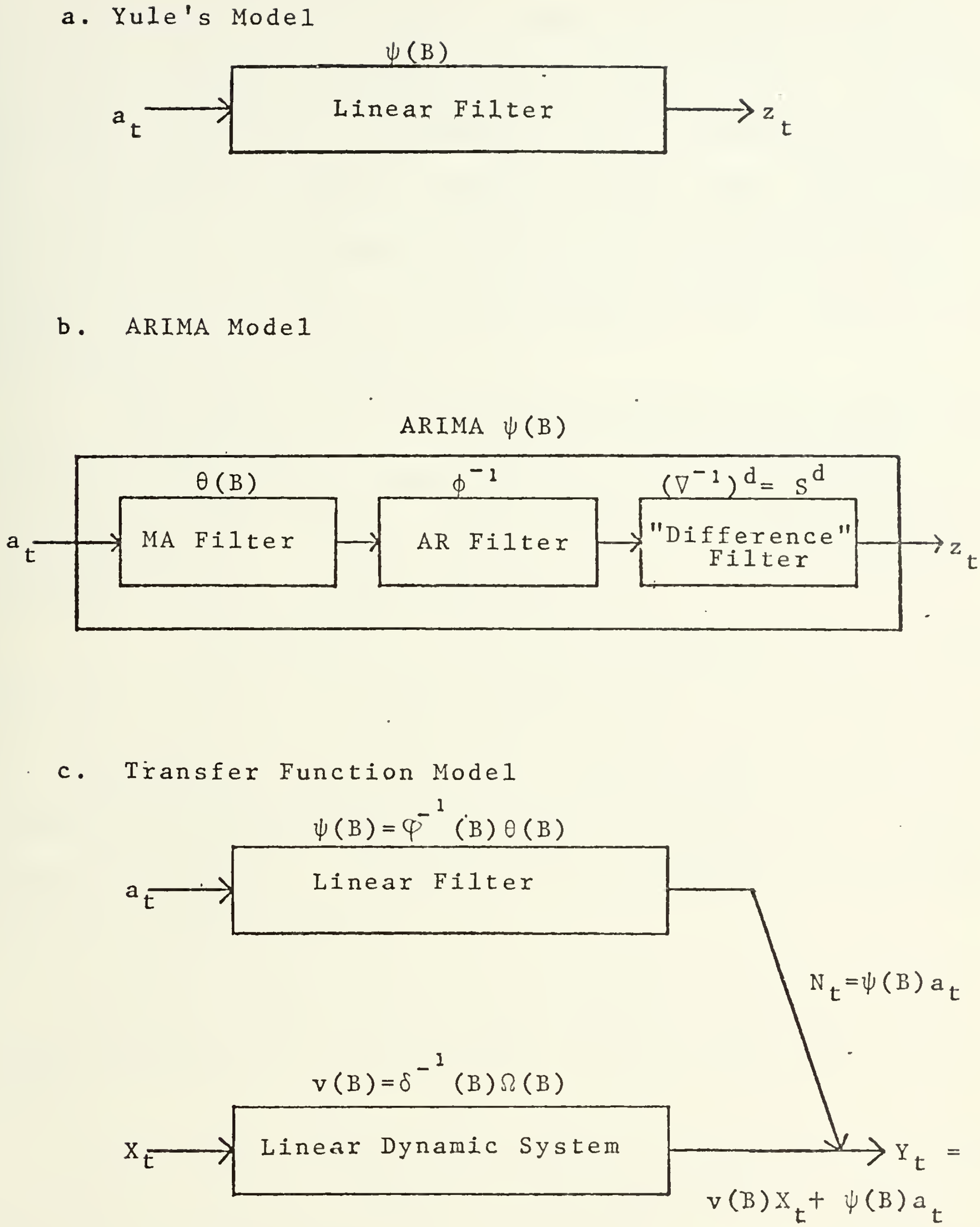


Figure 1. Filter Models
- after Box & Jenkins [1970]

just described. This class of models is known as the autoregressive integrated moving average (ARIMA) class. As the name suggests, this model incorporates three separate features, the autoregressive terms, moving average terms, and integration operator. The model is easier to understand if one discusses each part separately, building the total structure as new features are discussed.

The autoregressive model associates the current value of the process with a linear combination of past values of the process and a current random shock. An autoregressive process of order p, AR(p), would be:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \ldots + \phi_p \tilde{z}_{t-p} + a_t \qquad (2.3)$$

where $\tilde{z}_t = z_t - \mu$. If one considers the past values of $\tilde{z}_t$ as the "independent" variables, then one has the standard linear regression relation, hence the name autoregressive. The p+2 parameters: $\phi_1, \phi_2, \ldots, \phi_p, \mu, \sigma_a^2$, must generally be estimated directly from the data.

It is clear that by substituting back within the model (Equation 2.3) for the past values used by the model, e.g.

$$\tilde{z}_{t-1} = \phi_1 \tilde{z}_{t-2} + \ldots + \phi_p \tilde{z}_{t-p-1} + a_{t-1} \qquad (2.4)$$

one can eventually produce a representation of the AR(p) process as an infinite series in the a_t's, and hence, as a linear filter of the type described previously. Writing the Ar(p) process as

$$\phi_p(B) \tilde{z}_t = a_t \qquad (2.5)$$

and

$$\psi(B) a_t = \tilde{z}_t \tag{2.6}$$

then one obtains

$$\phi_p^{-1}(B) = \psi(B) \tag{2.7}$$

with the same constraints for stationarity of $\psi(B)$ as noted previously.

The second important process is the moving average process of order q, MA(q) which relates the current value of $\tilde{z}_t$ to past values of the a_t's. The MA(q) process may be represented as

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \tag{2.8}$$

Writing the MA(q) process as:

$$\tilde{z}_t = \theta_q(B) a_t , \tag{2.9}$$

it is immediately seen that this process already has the form of a linear filter where

$$\psi(B) = \theta_q(B). \tag{2.10}$$

In a formal statistical sense, the form described above is not confined to moving averages of the a_t's, since the θ_i's need not be positive, nor sum to one, but this is standard nomenclature nonetheless. The q+2 parameters of the model: $\theta_1, \theta_2, \ldots, \theta_q, \mu, \sigma_a^2$, again generally depend directly on the data for their estimation.

These two processes may be mixed to achieve parsimony of the model (discussed in a later section). The result is an autoregressive moving average process of order (p,q) that is, ARMA (p,q):

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \ldots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \quad (2.11)$$

or

$$\phi_p(B)\, z_t = \theta_q\, a_t. \quad (2.12)$$

Here a total of p+q+2 parameters: $\phi_1, \ldots, \phi_p$, $\theta_1, \ldots, \theta_q$, μ, σ_a^2, are required, and in general must all be estimated directly from the data. If q is finite, $\theta_q(B)$ is a stable filter. For the equivalent filter of the ARMA (p,q) process to be stable, then, it is sufficient to require $\phi_p^{-1}(B)$ to be stable, as the ARMA (p,q) equivalent filter is a simple infinite series:

$$\psi(B) = \phi_p^{-1}(B)\ \theta_q(B) \quad (2.13)$$

Further, it is clear that the ARMA (o,q) model is equivalent to the MA(q), and the ARMA (p,o) to the AR(p) when all are applied to the same $\tilde{z}_t$'s and a_t's.

For reasons of simplicity and stability of calculation, it is useful to require that the only ARMA(p,q) processes considered be those which are stable, and hence would result in a stationary z(t) if applied to a sequence of random shocks, a(t). But many observed time series are inadequately fitted by such an ARMA (p,q) model because they are non-stationary in nature. Many of these may be adequately fitted by using a generalized autoregressive operator $\varphi(B)$ in place of $\phi_p(B)$. The result will be an autoregressive integrated moving average model. This is shown schematically in Figure 1b.

It can be shown that the roots of the equation

$$
\begin{aligned}
0 &= \phi_p(B) \\
&= 1 + \phi_1 B + \phi_2 B^2 + \ldots + \phi_p B^p \qquad (2.14)
\end{aligned}
$$

where B is now considered in the same fashion as a variable in a polynomial, must lie outside the unit circle for $\phi_p^{-1}(B)$ to be stable. It can further be shown that if any of the roots lies inside the unit circle for either

$$0 = \phi_p(B) \tag{2.15}$$

or

$$0 = \theta_q(B) \tag{2.16}$$

then the ARMA(p,q) model exhibits explosive growth, thus violent non-stationarity. However, it turns out to be very useful in treating many types of non-stationarity to allow one or more of the roots of

$$\varphi(B) = 0 \tag{2.17}$$

to be equal to unity. Letting the ∇ operator symbolize this:

$$\nabla = (1 - B) \tag{2.18}$$

one can then write $\varphi(B)$ as:

$$
\begin{aligned}
\varphi(B) &= \phi_p(B)\, \nabla^d \\
&= \phi_p(B)(1-B)^d \qquad (2.19)
\end{aligned}
$$

where d is the number of roots of $\varphi(B)$ set equal to unity. Thus, the autoregressive integrated moving average model may be written:

$$\varphi(B)\tilde{z}_t = \theta_q(B)\, a_t \tag{2.20}$$

or

$$\phi_p\,(B)\, \nabla^d\, \tilde{z}_t = \theta_q(B)\, a_t. \tag{2.21}$$

This model can be symbolized by ARIMA (p,d,q). Substituting

$$w_t = \nabla^d \tilde{z}_t \tag{2.22}$$

produces a process in the w_t's which can be represented by a stable equivalent filter, since the process may be written

$$\phi_p(B)w_t = \theta_q(B)a_t \tag{2.23}$$

which is a simple ARMA (p,q) process in the w_t's. This reveals the types of non-stationarity which can be handled. In order for Equation (2.23) to be true, w(t) must be a stationary series. Hence, any series $\tilde{z}(t)$ which can be reduced to stationarity by forward differencing by the ∇^d operator can be adequately fit by such a model. This includes any series whose "trend" can be adequately approximated by a polynomial of degree d or less.

We note that the term integrated arises because

$$\nabla^{-1} z_t = S z_t = \sum_{h=-\infty}^{t} z_h \tag{2.24}$$

and, for example

$$S^3 z_t = \sum_{i=-\infty}^{t} \sum_{j=-\infty}^{t} \sum_{h=-\infty}^{t} z_h . \tag{2.25}$$

Thus, in forecasting a future value of z_t, say z_{t+1}, one uses the summation operator to go from the predicted and derived w_t's to the predicted z_{t+1}'s.

There is still one important characteristic that a time series may have that is useful to be able to handle: seasonality. Many treatments of seasonality are possible within the framework exhibited so far. This discussion will be limited to the multiplicative technique described below.

Box and Jenkins [1970] should be referenced for more detailed information on both this and other techniques.

The multiplicative technique is a simple extension of the ARIMA (p,d,q) model. The basic assumption is that the variation between observations separated by an increment of time equal to the period of the seasonal component can be adequately fitted by the same sort of models as used on adjacent observations. Hence, replacing B by B^s in the ARIMA (p,d,q) model, we get a seasonal representation:

$$\Phi_P(B^s)\ \nabla_s^D\ \tilde{z}_t = \Theta_Q(B^s)\ a_t \qquad (2.26)$$

or

$$W_t = (\Phi_1 B^{s \cdot 1} + \ldots + \Phi_P B^{s \cdot P})\ W_t$$
$$- (\Theta_1 B^{s \cdot 1} + \ldots + \Theta_Q\ B^{s \cdot Q})\ a_t + a_t \qquad (2.27)$$

or

$$W_t = \Phi_1 W_{t-s} + \ldots + \Phi_P W_{t-P \cdot s} + a_t$$
$$- \Theta_1 W_{t-s} - \ldots - \Theta_Q W_{t-Q \cdot s} \qquad (2.28)$$

where

$$W_t = \nabla_s^D\ \tilde{z}_t = (1 - B^s)^D\ \tilde{z}_t$$

But it is also clear that the process will also depend on adjacent values, so the complete multiplicative model, the ARIMA (p,d,q) x (P,D,Q) is:

$$\phi_p(B)\Phi_P(B^s)\nabla_s^D\ \tilde{z}_t = \theta_q(B)\ \Theta_Q(B^s)a_t \qquad (2.29)$$

This result is a very powerful class of models for fitting naturally occurring time series. This is the class upon which parametric time series analysis is based.

3. Transfer Function Models

Box and Jenkins [1970] further discuss transfer function models designed to fit cases in which two processes are related by

$$(1+\Xi_1 D+\ldots+\Xi_R D^R)\ Y(t) = (H_0+H_1 D+\ldots+H_s D^s)X(t-\tau) \qquad (2.30)$$

where $D \equiv d/dt$, the Ξ's and H's are unknown parameters, and τ is the dead-time or delay in the system. Where Y_t and X_t are discrete time series measured at equispaced times, (2.30) becomes

$$(1+\xi_1\nabla+\ldots+\xi_r\nabla^r)Y_t = (\eta_0+\eta_1\nabla+\ldots+\eta_s\nabla^s)X_{t-b} \qquad (2.31)$$

where D is replaced by $\nabla \equiv 1-B$ as before. Substituting $B = 1-\nabla$, we obtain

$$(1-\delta_1 B-\ldots-\delta_r B^r)Y_t = (\omega_0-\omega_1 B-\ldots-\omega_s B^s)X_{t-b}$$

or

$$\delta(B)Y_t = \omega(B)B^b X_t$$

$$= \Omega(B)X_t \qquad (2.32)$$

As in the ARIMA (p,d,q) case, this relation may be written in the form of a linear filter:

$$Y_t = \delta^{-1}(B)\Omega(B)X_t = v(B)X_t \qquad (2.33)$$

which is stable if the v(B) series converge for B given a numerical value such that $0 < |B| \leq 1$. It is to be noted that the transfer function model and the stochastic ARIMA models are somewhat parallel.

Now, remembering the discussion on the lack of deterministic time series, it is necessary to consider a stochastic transfer function model by adding superimposed noise, N_t,

to the process, where N_t and X_t are independent. Hence, we obtain

$$Y_t = v(B)X_t + N_t \tag{2.34}$$

for the stochastic model. But there is no reason to suppose that the process does not effect the noise in some way, or that the noise contribution is unstructured. Therefore, the idea is advanced that the noise, N_t, will be represented as having the form

$$N_t = \psi(B)a_t = \varphi^{-1}(B)\theta(B)a_t \tag{2.35}$$

where the ψ's, φ's, and θ's are as before. This form includes unstructured noise, the raw random shocks, a_t's, as a single subset of the possible structures. Hence, the model one would be led to fit is

$$\begin{aligned} Y_t &= v(B)X_t + \psi(B)a_t \\ &= \delta^{-1}(B)\Omega(B)X_t + \varphi^{-1}(B)\theta(B)a_t \end{aligned} \tag{2.36}$$

This is shown schematically in Figure 1c.

E. APPLICATIONS TO WIND-WAVE COUPLING

For an application of these techniques to the problem of detection of hypothesized organized motion in air adjacent to natural waves, it seems clear that a transfer function stochastic model is likely to give the most direct information as to the structure of the relationship between the waves and the adjacent airflow. One would fit the transfer function model by regarding the wave heights as inputs, X_t, and a wind velocity component as the output, Y_t. The existence of significant values for the v's would immediately suggest that there

are fluctuations in the overlying airflow which are not random in relation to the waves, particularly if such values appear repeatedly for data from different observational periods. It is a fortunate circumstance that statistical tests of significance for these parameters exist. However, the accuracy of these tests depends on how closely the parent distribution of the actually calculated random shocks corresponds to the distribution picked to be the theoretical parent of the random shocks in the model. The structure of the fitted model may give clues as to the nature of the actual process. It is an unfortunate circumstance, however, that the transfer function models are beyond the scope of this paper.

The analyses performed were involved fitting ARIMA (p,d,q) x (P,D,Q)'s models to individual time series. Three benefits may be possible from this procedure. First, models fitted individually to wind components and wave heights from the same observations may be compared for suggestions of interaction. This type comparison is difficult, however, because many individual models with differing orders in the various types of processes may be quite similar in result and fit, depending on the data. Hence, for example, one may not be able to say that an ARIMA (1,d,0) x (0,0,0) represents data with a different structure than that of a series best fit by an ARIMA (0,d,2) x (0,0,0). The latter arises because these two models closely correspond for certain sets of values of the parameters involved.

The next benefit is that the fitted model may suggest in itself a possible structure for the process. Also possible and more likely is the fact that repeated analyses of this kind may help a researcher following an adaptive model building strategy to infer a general model for given conditions. From this, he may infer some of the structure of the parent process. The researcher is helped in this quest by the statistical confidence intervals it is possible to obtain for parameter estimates and joint confidence intervals for groups of parameters (with the same caution here as presented in the transfer function model discussion directly above).

Third, and seemingly most important, this study may give an idea of the usefulness of these techniques in fitting to boundary layer data. The study intended to give an indication whether such data is of the types that this sort of analysis can handle, for pre-evaluation of the usefulness of further studies along these lines.

A final note of interest is that Yule's proposition [Yule, 1927], discussed in I-D, may have a physical significance for wind-wave data. It may be possible to view the generalized turbulence of the airflow adjacent to natural waves as a parent random process producing a series of a_t's, and the waves as acting like a linear filter on these a_t's to produce organized motion and a resultant structuring of the noise elements, that is, the observed turbulence. Recent papers by Reynolds [1968] and Davidson and Frank [1972] suggest that the structure of the turbulent field is important in many

facets of air-sea interaction. Hence, parametric time series analysis may hold great potential in boundary layer studies. However, to reiterate a previous statement, it is not necessary that there exist any correspondence at all between the actual generating process and the Yule viewpoint for this type of analysis to be beneficial.

III. ANALYSIS CONSIDERATIONS

The data considered were from an August 19 observation period, described in Davidson [1970]. From these data three time series were chosen: wave height, u-component of the wind 1.5 meters above surface, and w-component of the wind at 1.5 meters. ARIMA models were fitted to each of these series, and examined for goodness of fit. The models were examined for information contained in them about the structure of the time series to which they were fit.

The primary feature of parametric time series analysis is model fitting. An iterative approach to model building was used which allows the data to influence the model structure. The steps in this procedure are discussed in Section A. Again, a more complete discussion of this procedure is found in Box and Jenkins' text [Box and Jenkins, 1970, p. 18].

A. MODEL BUILDING

In fitting ARIMA models, an iterative model building procedure is most fruitful. The steps of such a procedure are described below and the process is summarized in Figure 2.

1. Selection of General Class of Model

The first step in model building is to postulate a general class of models to be fitted. By choosing to perform parametric time series analysis, the researcher has implicitly chosen to use the ARIMA class of models. However, further choices are possible at this stage. Due to the very nature

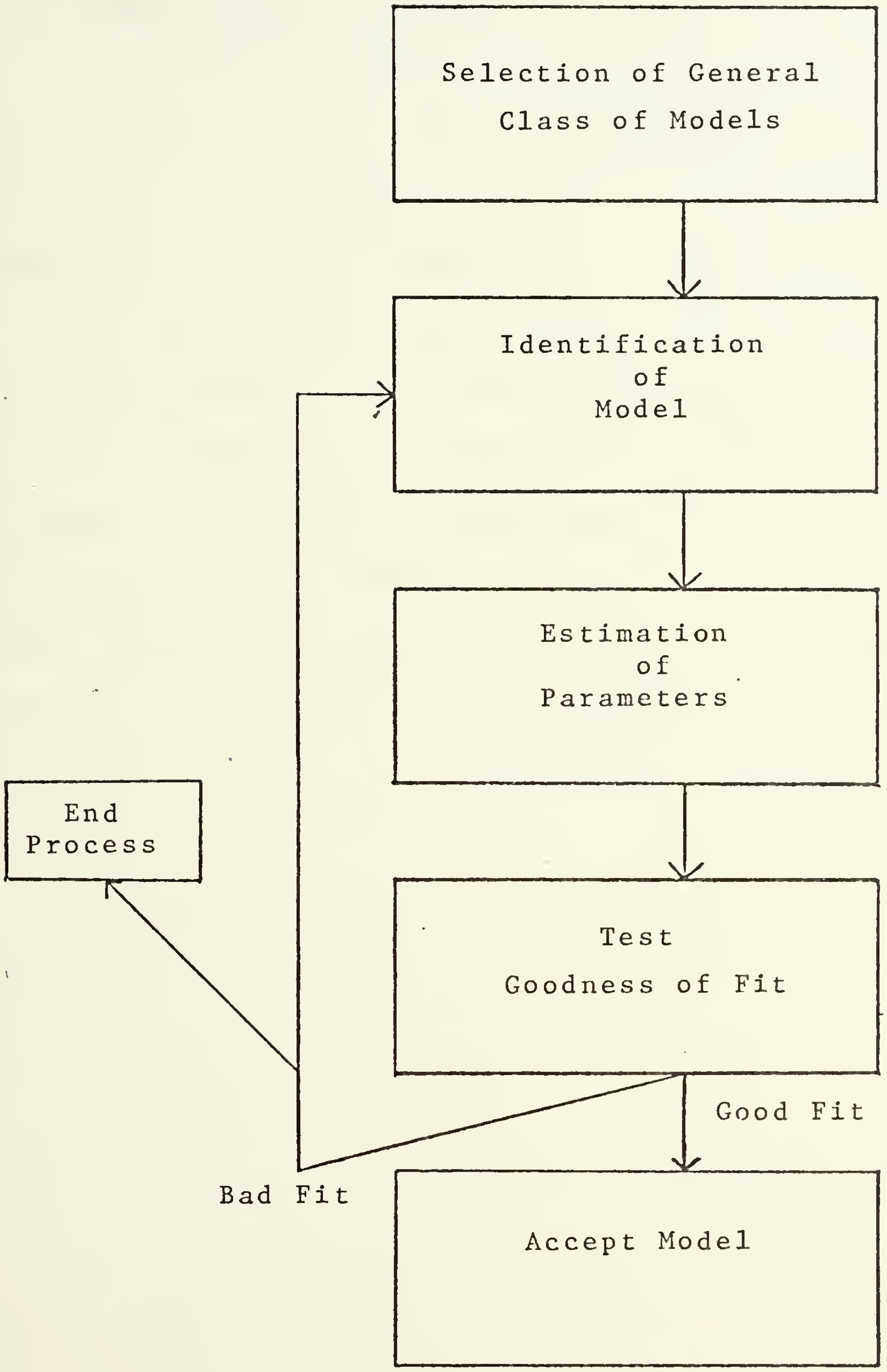


Figure 2. Model Building
- after Box & Jenkins [1970]

of the problem, one is virtually forced to consider the seasonal extension to the ARIMA (p,d,q) model.

2. Identification of Model

The second step in the procedure is the identification of a particular model or subset of models from the general class chosen. This process can be influenced both by outside knowledge and information contained in the data.

Having chosen to consider an ARIMA (p,d,q) x (P,D,Q) model, theory suggests that the order parameter d should be zero, and D should not. The need for $d \neq 0$ would imply a noticeable change in the mean sea level was taking place during the period of observation. The need for D=0 would imply that no seasonal change in the time series at all was taking place during the period of observation. Seasonal variation certainly does occur in the waves, and we expect seasonal variation in the air.

It can be shown that the stable non-stationary operator $\nabla_s = (1-B^s)$ has zeroes at $e^{\pm ik2\pi/s}$ $(k=0,1,\ldots,s-1)$. The application of this operator to a time series will remove a seasonal component of the form

$$\text{seasonal component} = b_o + \sum_{j=1}^{[s/2]} \{b_{1j}\cos(\frac{2\pi j}{s}) + b_{2j}\sin(\frac{2\pi j}{s})\}$$

where the b's are adaptive coefficients implicit in the data, and $[S/2] = S/2$ if S is even, or $[S/2] = 1/2\ (S-1)$ if S is odd. This type of operator is a parsimonious representation of seasonal components of a type which require many sine and cosine waves to represent them adequately. An example of

such a seasonal component would be a spike at periodic intervals.

Theory predicts that the main seasonal effect will have the form of a single sine wave at the frequency of the water waves. The ∇_s^D operator is not a parsimonious representation of a single sine wave. Furthermore, the adaptive coefficients of at least some of the additional sine and cosine waves contained in the ∇_s^D operator are likely to be inflated by intermittent and statistically non-significant periodicities in the random turbulent component of the motion. Therefore, an operator representing a single sine wave at the desired frequency, adaptive in phase and amplitude to the data, was applied in place of the ∇_s^D operator in some models. This operator is $[1-2\cos(2\pi/p)B + B^2]$, where p is the desired period. This stable non-stationary operator, which has zeroes at $e^{\pm i2\pi/p}$, will remove a seasonal component of the type:

$$\begin{aligned} \text{seasonal component} &= b_1 \sin(\frac{2\pi}{p}) + b_2 \cos(\frac{2\pi}{p}) \\ &= b_3 \sin(\frac{2\pi}{p} + \delta) \end{aligned}$$

where the b's are adaptive coefficients, δ is an adaptive phase angle, and p is the desired period.

Parsimony is a primary consideration at this stage. Parsimony involves making use of the most efficient type of model possible. For example, while an AR(1) process can be theoretically represented by an infinite MA process, and by a finite MA(q), q > 1, process in practice, it is most parsimoniously represented by the AR(1) model involving only one

parameter as opposed to an MA(q) process involving $q > 1$ parameters. Similarly, an AR(p) model is not a parsimonious representation of an MA(1) process, nor is either an AR or MA model a parsimonious representation of a mixed ARMA process.

Hence, it is necessary to go to the data for an indication of the values of p,q,P, and Q which are best fit, remembering that in practice a value of two or less for each is normally sufficient. To get this indication, one examines the autocorrelation function and partial autocorrelation function. A brief discussion of what one looks for is to be found in Appendix A. Estimates of p,q,P and Q are made, and rough estimates of the fitted parameters, the ϕ's, Φ's, θ's, and Θ's, are made.

3. Estimation

The third stage of the procedure is to fit the model to the data. The rough estimates of the fitted parameters are now used as initial guesses for iterative estimation routines. At this stage, the best fit for the model chosen is obtained.

4. Diagnostic Checks

The final stage is diagnostic checking to determine if the fitted model is adequate. If it is, the procedure ends, and the model is chosen. If the model is determined to be inadequate or unsatisfactory, the process is iterated from the second step until a suitable model is either found or determined to be non-existent.

A more complete discussion of the diagnostic tools used is to be found in Appendix B.

B. PRELIMINARY ANALYSES

The preliminary analysis produced estimated autocorrelation function values, estimated autocovariance function values, and estimated partial autocorrelation function values of the time series which resulted from the application of the $\nabla^d \nabla_s^D$ operator, with various values of d and D, and appropriate values of s, to the initial data. These estimated function values were also calculated for the time series resulting when ∇_s^D was replaced by $[1-2 \cos (2\pi/p)B + B^2]$, the less general operator discussed in A-2. From these estimated function values, the proper types and degrees of differencing needed to produce a stationary time series were estimated. Further analysis of the estimated function values for the properly differenced series resulted in estimates for the proper values of the parameters p,q,P, and Q, for the multiplicative ARIMA (p,d,q) x (P,D,Q) model to be fit. The set of the most likely model candidates was the result.

C. ANALYSES

The preliminary analysis yielded the set of the most likely model candidates; the parameters required by these models, the ϕ's, Φ's, etc., were estimated by mapping the maximum likelihood surface of the parameters, as reflected by the sum of the squares of the estimated a_t's (that is, the sum of the squares of the residuals). Thus, these estimates were "least squares" estimates, and this method is an accepted routine for non-linear least squares estimation. Diagnostic checks

were then applied to the a_t's resulting from the use of the best estimates for the parameters in each model. These diagnostic checks included comparing the magnitudes of the sums of the squares of the residuals of different models, calculating the autocorrelations of the residual series, and calculating a chi-square test for adequacy of fit [Box and Jenkins, 1970, p. 291]. These diagnostic checks were then compared in an effort to determine the best model of those examined.

Unfortunately, the author was unable to get a Marquardt maximum descent nonlinear least squares algorithm to work. Had this been done, reasonably accurate confidence intervals and joint confidence intervals for parameter estimates would have been easily available in addition to the speedier arrival at the least squares estimates. As this was not done, alternate methods for obtaining these estimates, set forth in Box and Jenkins [1970, p. 228], were explored. Considering the lack of real need for such interval estimates in this study, due to the small number of time series analyzed and subsequent preclusion of a meaningful attempt at an adaptive model building strategy, these methods were found to be too inefficient and time consuming to be of sufficient relative value to attempt. Hence, although no confidence intervals were estimated, their exclusion was not significant in this study.

In summary, the estimated autocorrelation functions, estimated autocovariance functions, and estimated partial autocorrelation functions of the time series resulting from

application of the difference operators to the original time series were used to identify the proper degree of differencing and the most likely types of models to be fit to the properly differenced series. The parameters of these models were estimated by a nonlinear least squares mapping technique. The residuals, the a_t's resulting from each type of model when the least squares estimates for the model's parameters were used, were tested for indications of the model's goodness of fit. The results of these tests were compared, and the best model was chosen.

IV. RESULTS

The results of the parametric time series analyses were in good agreement with the spectral analyses done with the same data by Davidson [1970]. The results clearly indicate the tractability of this type of data to this type of analysis.

A. PERIODIC EFFECTS

The $\nabla^d \nabla_s^D$ difference operator was applied to the data for several values of d and D. Table I contains the values applied. From both spectral analysis and autocorrelation function estimates, it was determined that the dominant period of the waves was best estimated as six seconds, corresponding to a period of 30 lags in the time series, while the air was observed to have a dominant period of only 4.7 seconds, corresponding to 24 lags in the time series, in both the horizontal and vertical components. The estimated autocorrelation functions of all three time series before differencing, Figure 3, shows the form of a damped sine wave, but the damping is not great. The autocorrelations at high lag numbers do not fall off sufficiently fast to support an assumption that these autocorrelation functions have arisen from stationary time series. It can be seen in Figure 4 that application of the ∇_s^D operator did not result in a picture more compatible with the assumption that stationarity resulted. None of the differencings involving the $\nabla^d \nabla_s^D$ operator were satisfactory

TABLE I

DEGREES OF DIFFERENCING EXAMINED

$$\nabla^{d}\nabla_{s}^{D}z_{t} \quad \text{or} \quad [1-2\cos(2\pi/p)B+B^{2}]z_{t}$$

p=s=24 for Airflow; p=s=30 for Wave heights

run no.	d	D	S	$[1-2\cos(2\pi/p)B+B^{2}]$
1	0	0	No	No
2	1	0	No	No
3	2	0	No	No
4	0	1	24/30	No
5	0	2	24/30	No
6	1	1	24/30	No
7	1	2	24/30	No
8	2	1	24/30	No
9	2	2	24/30	No
10	0	0	No	Yes
11	1	0	No	Yes
12	2	0	No	Yes
13	0	1	4	Yes

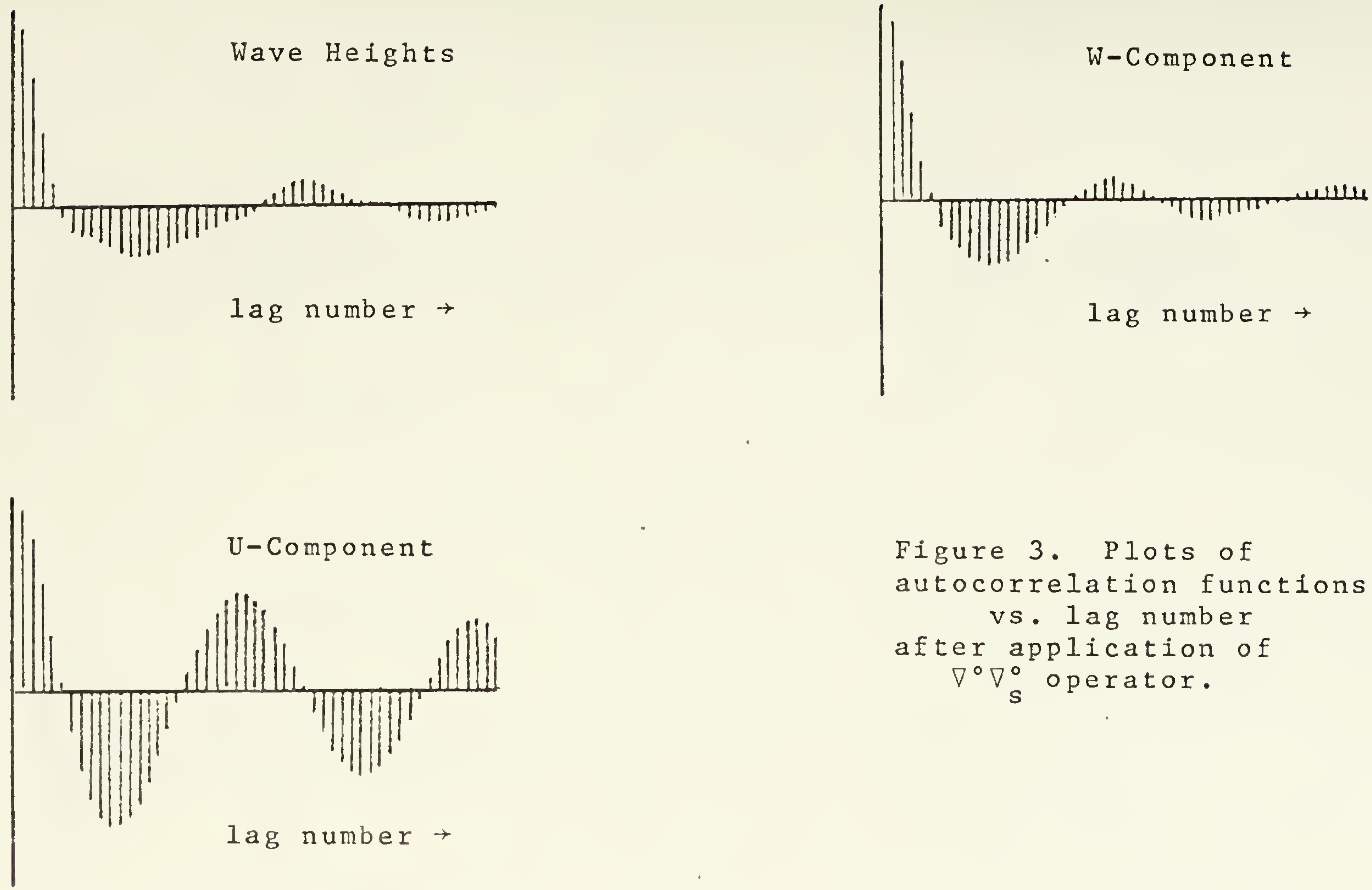
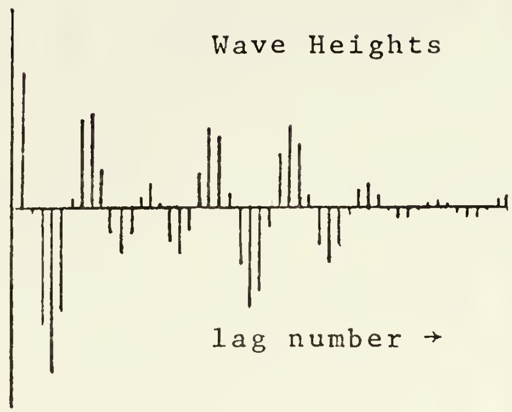
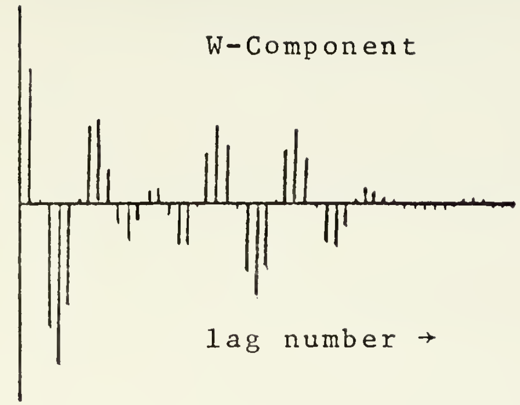


Figure 3. Plots of autocorrelation functions vs. lag number after application of $\nabla^\circ\nabla_s^\circ$ operator.





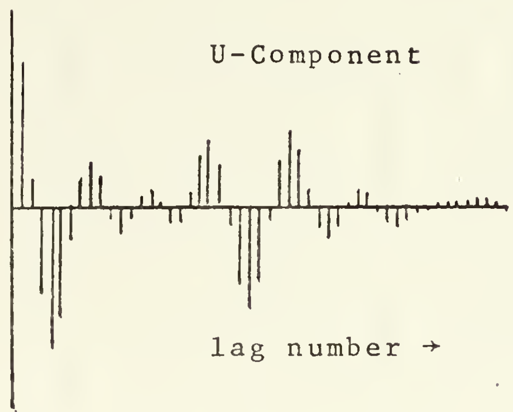
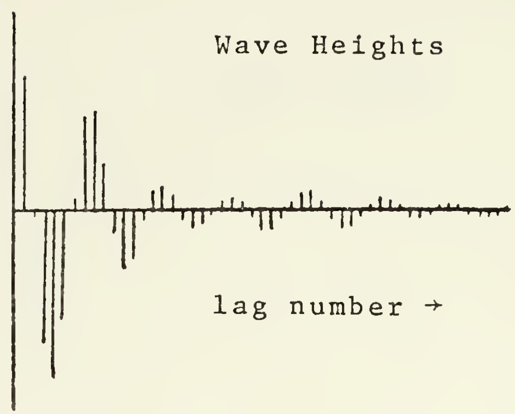


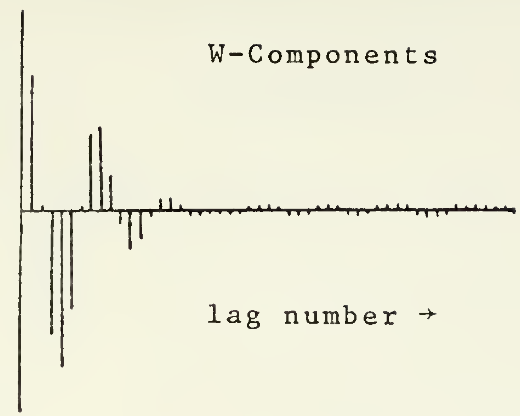
Figure 4. Plots of autocorrelation functions vs. lag numbers after application of $\nabla^{\circ}\nabla^{1}_{30}$ operator on the wave series and $\nabla^{\circ}\nabla^{1}_{24}$ operator on the airflow series.

when $D \neq 0$. Applying the ∇^d alone, with $d = 2$, resulted in a satisfactory pattern for this data (Figure 5). However, the difference between the ∇^2 operator and the non-general operator $[1-2 \cos (2\pi/p)B + B^2]$, discussed in III-A-2, is small for p=24 for the wind, and p=30 for the waves. The $[1-2 \cos (2\pi/p)+ B^2]$ operator gives a slightly better result (Figure 6) and with the support from theory, the assumption that this operator is the proper one seems justified. No value $d \neq 0$ gave a better result than d=0 when the full $\nabla^d[1-2\cos(2\pi/p)B + B^2]$ operator was applied, so the final series considered was

$$w_t = [1-2 \cos (2\pi/p) B + B^2] z_t$$

The result that no local differencing was necessary (that is, for ∇^d, d=0), and that the best results arose from the application of the non-general operator which took out a single sine wave, is precisely in agreement with the hypothesis concerning the organized motion considered in this paper. The fact that the frequency of the organized motion in the air is not the same as the dominant frequency of the waves is not in good agreement with the hypothesis considered, though it is in agreement with Davidson's spectral results [Davidson, 1970]. The lack of efficiency in the ∇_s^D operator for production of stationarity was judged to be due to spurious periodicities in the random turbulent components of the series.





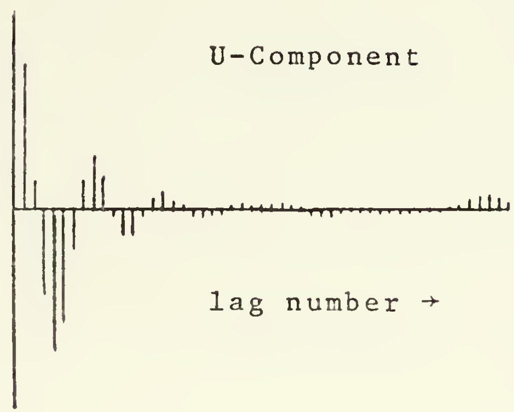
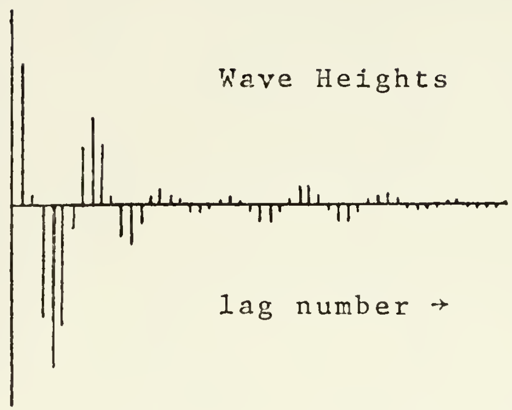
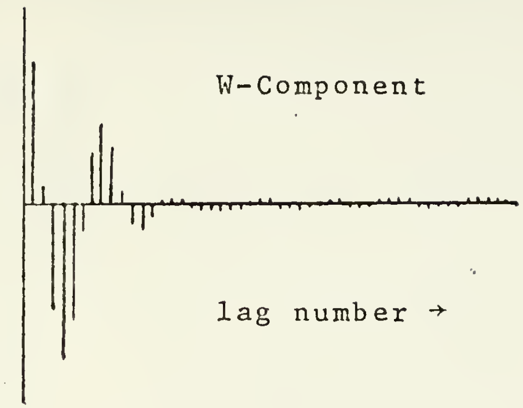


Figure 5. Plots of autocorrelation functions vs. lag number after application of $\nabla^2\nabla_s^\circ$ operator.





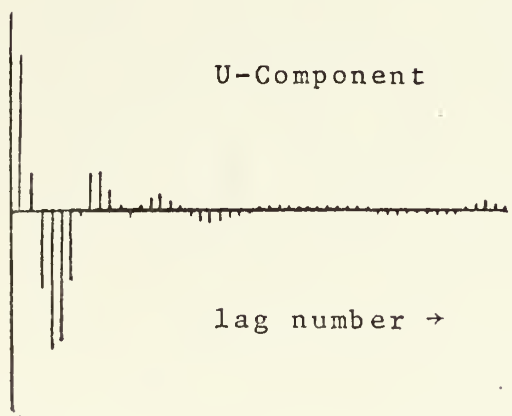


Figure 6. Plots of autocorrelation functions vs. lag number after application of the $[1-2\cos(2\pi/30)B+B^2]$ operator to wave series and the $[1-2\cos(2\pi/24)B+B^2]$ operator to the airflow series.

B. ARMA MODEL FITTING

Once a satisfactorily stationary time series was produced, a closer examination of the estimated autocorrelation and estimated partial autocorrelation functions revealed that these functions showed unmistakable evidence that a mixed model, that is, one requiring both autoregressive and moving average terms, was required for each series. However, the three series each showed different structure, and were considered separately.

1. Wave Model

The estimated autocorrelation function of the waves was most easily diagnosed for indications of the proper model. The pattern exhibited is that of an ARMA (p,q) x (P,Q) where p=0, q=1, P=1, Q=1. The pattern was quite distinct, and overfitting, by letting p=1, q=2, P=2, q=2, resulted in estimates for the ϕ_1, Φ_2, θ_2, and Θ_2 parameters which were close to zero, while no significant decrease in the sums of the squares of the residuals was achieved. Fitting a smaller model, one with fewer parameters, resulted in a significant increase in the sum of the squares of the residuals. Hence, an ARMA (0,1) x (1,1) was considered best, and the final fitted model was

$$(1-\Phi_1 B^{30})w_t = (1-\theta_1 B)(1-\Theta_1 B^{30})a_t$$

where $\Phi_1 = +0.11$, $\theta_1 = -0.53$, and $\Theta_1 = -0.33$.

2. U-Component Model

The estimated autocorrelation function of the u-component series did not immediately correspond to a particular

model in the way that the waves' function did. No particular need for a seasonal component to the model, other than differencing, was suggested. Hence, P and Q were initially taken as zero, and subsequent fitting of ARMA (p,q) models showed that the best model of this type was the ARMA (1,1). The final fitted model of this type was, after overfitting in both local and seasonal parameters,

$$(1-\phi_1 B)w_t = (1-\theta_1 B)a_t$$

where $\phi_1 = -0.31$, and $\theta_1 = -0.74$.

It was interesting to note that the estimated partial autocorrelation showed large values at periodic intervals. There seemed to be two periods involved, one of 24 and the other of 31. This is a possible indication that a seasonal autoregressive operator might be useful. Two such models were tried, one with s=24 and one with s=31:

$$(1-\phi_1 B)(1-\Phi_1 B^{31})w_t = (1-\theta_1 B)a_t$$

and

$$(1-\phi_1 B)(1-\Phi_1 B^{24})w_t = (1-\theta_1 B)a_t$$

Each of these models produced improvements in the diagnostic checks of the a_t's, but the improvement was slight enough to make it questionable if these are parsimonious operators.

3. W-Component Model

The estimated autocorrelation function and the estimated partial autocorrelation function of the w-component series showed no indication that seasonal parameters other than differencing were required. After overfitting in both

local and seasonal parameters, the ARMA (1,1) x (0,0) was judged best. The final model was

$$(1-\phi_1 B)w_t = (1-\theta_1 B)\ a_t$$

where $\phi_1 = -0.09$ and $\theta_1 = -0.87$.

C. SUMMARY

All three time series were most adequately reduced to stationarity by application of a difference operator which removed a single sine wave, with adaptive amplitude and phase angle, from the data. The period of the motion in the air thus removed differed from that of the waves, the period in the air being 4.7 seconds, as opposed to the 6.0 second dominant period in the waves.

Once reduced to stationarity, the wave series was best represented by an ARMA (0,1) x (1,1) model The stationary series from the u-component was adequately represented by either an ARMA (1,1) x (1,0) or an ARMA (1,1) model. The w-component was best represented by an ARMA (1,1) model.

The data proved to be of the type for which parametric time series analysis is useful. The results show good agreement with previous results obtained by spectral analysis of the same data by Davidson [1970].

V. CONCLUSIONS

The necessity for removing a single sine wave from the data to achieve stability strongly suggests that there was a significant sinusoidal component of organized motion in the air flow immediately over the waves during the period in which the data were gathered. That this sinusoidal motion has a dominant period which is different than that of the waves is not supportive of the theories which predict that the air will have organized motion with the same period as the waves. However, it is noted that the periods in question differ only by 1.3 seconds, although this appears significant in this densely sampled time series, which had about 25 observations per period (observations every .2 seconds).

The fact that none of the series reduced to a random walk model, that is,

$$Z_t = a_t,$$

after being differenced to stationarity, suggests that either further organized motion exists, or that some sort of structure is imposed upon the turbulence. Due to the fact that only one series of each type was considered, the significance of the ARMA models which were fitted to the properly differenced series is unknown. Too few series were considered to be able to generalize about the structure of the turbulent flow from the ARMA models derived here. The possibility clearly exists that the repeated fitting of such models may

result in a formulation of a general model by an adaptive strategy.

The most important conclusion of this exploratory work is that parametric time series analysis may be applied to wind-wave data with a reasonable expectation that results of value can be obtained. It is noted in this respect that this analysis closely agreed with the traditional spectral approach in their region of overlap. The results in this study show that the examination of the estimated autocorrelation and partial autocorrelation functions of various degrees of differencing of the original series may be of value in understanding the structure of such series, even if the actual fitting of the ARMA model to the resulting series is not attempted. In short, the parametric approach appears to hold considerable potential in the analysis of organized motion within turbulent boundary layer data.

APPENDIX A

Model Identification Considerations

The initial identification of the proper form of the ARIMA to best fit the data is based on examination of the estimated autocorrelation function and the estimated partial autocorrelation function of the series. There are two distinct steps involved: identifying the degree of differencing necessary and identifying the resultant ARMA model. These steps are described briefly below; they are more fully developed in Box and Jenkins [1970, pp. 174-207; 300-334].

The first step in initial model identification is determining the degree of differencing necessary to produce a stationary time series from the raw data. Thus one is attempting to estimate the proper values of D and d in the $\nabla^d \nabla_s^D$ operator. The distinctive characteristic of a non-stationary time series for identification purposes is the fact that the autocorrelation function does not go quickly to zero with increasing lag number. The autocorrelations may decrease with lag number, but the decrease will be closer to linear than the exponential damping expected with a stationary time series. Further, recurrent patterns spaced a fixed distance in time (at a fixed number of lags apart) may indicate the need for seasonal differencing, if the autocorrelation coefficients do not fall off sufficiently fast. Hence, one examines first the estimated autocorrelations of the original data, and then successive degrees of differencing, both

seasonal and local, until a pattern consistent with a stationary series is found.

The second step is to examine the autocorrelations arising from the properly differenced series for indications of the resulting ARMA model to fit to the differenced series. One is aided in this by the knowledge of the behavior of the autocorrelation and partial autocorrelation functions for various models. The theoretical autocorrelation function of an AR(p) model tails off and the partial autocorrelation function abruptly goes to zero after p lags. The theoretical autocorrelation function of a MA(q) process goes abruptly to zero after q lags, while the partial autocorrelation function tails off. Both functions tail off in a mixed model. The estimated functions will generally follow the theoretical functions. However, the estimates have large variances, and the match need not be close. In general, it is possible to produce an autocovariance generating function for any given model. This function can be used to show the form of the theoretical autocorrelation for any model. One can then compare the estimated autocovariance (or autocorrelation) function to the theoretical possibilities to find a reasonable match.

In summary, the estimated autocorrelation and partial autocorrelation functions are examined for characteristic patterns which indicate the proper degree of differencing and the proper ARMA model (to be fit to the properly differenced series). Due to the variability of such estimated

functions, precise definition of model is often impossible, and the skill and experience of the analyst may be significant in obtaining the best results.

APPENDIX B

Diagnostic Checks of Model Adequacy

In order to insure that the model which has been fit is adequate, it is useful to have available diagnostic procedures for testing the adequacy of a given model. Such tests may be based on both evaluations at the total model level and at the level of the residuals alone. Several diagnostic tests which were used are discussed below; these tests, and others, are more fully described in Box and Jenkins [1970, pp. 287 ff].

A primary diagnostic check on the total model level is the comparison of the magnitudes of the sums of the squares of the residuals between models. A significant decrease in this value with a change in model indicates that the original model was not adequate. A second check on the whole model level is overfitting of parameters. If one judges, for example, that an ARIMA (p,d,q) is adequate while noting that if it were inadequate, it would likely be inadequate in the AR parameter, then one might fit an ARIMA (p+1,d,q) model. One could then compare results of original model and the overfitted model for indications of the original model's adequacy. That is, if no significant change in fit occurred, as would be the case if the best estimates of the extra parameters in the overfit model were all close to zero, then one is reassured as to the adequacy of the model.

To overcome the need for the prior knowledge of the suspected deficiencies in the model, such prior knowledge being

implied in overfitting, it is possible to use diagnostic checks on the residuals generated by the model. The autocorrelation function, or equivalently the spectrum, of the residuals can be examined for indications that the residuals generated do not approximate white noise. Individual autocorrelations may be tested for significant deviation from a value of zero by Student-t tests, using the appropriate limits for estimated residuals rather than those for white noise. Finally, the first K autocorrelations of the residuals may be looked at as a whole. The value Q such that

$$Q = N \sum_{m=1}^{K} r_m^2 (a),$$

where $r_m^2 (a)$ is the residual autocorrelation coefficient at lag m, and N = number of observations less d, the differencing parameter, is distributed as $\chi^2(K-p-q-P-Q-M)$. Here, the p,q,P, and Q are as before, and M=1 if the mean were removed from the initial series before the analysis, and M=0 otherwise. Hence, the final test discussed here is an χ^2-goodness of fit test.

In summary, tests may be made either between models or within models to determine the adequacy of models. Both types of tests were used in the analysis presented in this paper.

BIBLIOGRAPHY

1. Birnbaum, Z. W., 1962: Introduction to Probability and Mathematical Statistics, New York, Harper and Brothers Publishers, p. 225-242.

2. Blackman, R. B., and Tukey, J. W., 1958: The Measurement of Power Spectra, New York, Dover Publications, Inc., 190 pp.

3. Box, G. E. P., and Jenkins, G. M., 1970: Time Series Analysis Forecasting and Control, San Francisco, Holden-Day, Inc., 553 pp.

4. Jenkins, G. M., and Watts, D. G., 1968: Spectral Analysis and Its Applications, San Francisco, Holden-Day, Inc., 525 pp.

5. Kendall, M. G., and Stuart, A., 1966: The Advanced Theory of Statistics, Volume 3, New York, Hafner Publishing Co., p. 342-503.

6. Davidson, K. L., and Frank, A. J., 1972: "Wave-Related Fluctuation in the Airflow above Natural Waves," submitted to J. Phys. Ocean., June 1972.

7. Holland, J. Z., 1972: "A Statistical Method for Analyzing Wave Shapes and Phase Relationships of Fluctuating Geophysical Variables," J. Phys. Ocean., 2.

8. Miles, J. W., 1957: "On the Generation of Surface Waves by Shear Flows," Part 2, J. Fluid Mech., 6(4), p. 568-582.

9. Yule, G. U., 1927: "On a Method of Investigating Periodicities in Disturbed Series with Special Reference to Wölfer's Sunspot Numbers," Phil. Trans., A226, p. 267.

10. Reynolds, W. C., 1968: The Mechanics of an Organized Wave in a Turbulent Shear Flow, Unpublished Manuscript, Department of Mechanical Engineering, Stanford University, 39 pp.

11. Frank, A. J., 1971: An Investigation of the Properties and Influence of Wave-Induced Organized Motion in the Adjacent Airflow, Master's Thesis, Naval Postgraduate School, 84 pp.

12. Sokol, S. E., 1971: Use of Correlation Techniques in Automatic Detection of Thermal Plumes in the Air Layer Adjacent to Natural Water Waves, Master's Thesis, Naval Postgraduate School, 41 pp.

13. Davidson, K. L., 1970: An Investigation of the Influence of Water Waves on the Adjacent Airflow, ONR Contract No. N00014-67-A-0181-0005, University of Michigan, 239 pp.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Professor Kenneth L. Davidson, Code 51Ds Department of Meteorology Naval Postgraduate School Monterey, California 93940	5
4.	Professor Noel Boston, Code 58Bb Department of Oceanography Naval Postgraduate School Monterey, California 93940	1
5.	Ensign Craig H. Smith, USN 4014 Shore Drive Virginia Beach, Virginia 23455	3
6.	Professor Edward Thornton, Code 58Tm Department of Oceanography Naval Postgraduate School Monterey, California 93940	1
7.	Department of Meteorology, Code 51 Naval Postgraduate School Monterey, California 93940	2
8.	Professor G. J. Haltiner, Code 51Ha Department of Meteorology Naval Postgraduate School Monterey, California 93940	1
9.	Professor R. L. Alberty, Code 51Al Naval Postgraduate School Monterey, California 93940	1
10.	Professor R. L. Elsberry, Code 51Es Department of Meteorology Naval Postgraduate School Monterey, California 93940	1

11. Dr. Donald J. Portman 1
Department of Meteorology and Oceanography
University of Michigan
Ann Arbor, Michigan 48103

12. Professor R. T. Williams, Code 51Wu 1
Department of Meteorology
Naval Postgraduate School
Monterey, California 93940

13. Naval Weather Service Command 1
Washington Navy Yard
Washington, D. C. 20390

14. Assistant Professor F. R. Richards 1
Code 55Rh
Naval Postgraduate School
Monterey, California 93940

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

ORIGINATING ACTIVITY (Corporate author) Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

REPORT TITLE

Application of Parametric Time Series Analysis Methods to Descriptions of Wave-Induced Fluctuations in the Adjacent Airflow

DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Master's Thesis; September 1972

AUTHOR(S) (First name, middle initial, last name)

Craig H. Smith

REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1972	62	13

8a. CONTRACT OR GRANT NO. b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

0. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School Monterey, California 93940

3. ABSTRACT

Certain methods of analysis for testing the hypothesis that organized motion exists in the airflow above naturally occurring water waves are examined. In particular, two current methods, spectral analysis and joint probability density function analysis, are briefly discussed; two new methods, matched filters and parametric time series analysis, are suggested.

An analysis is done with parametric time series analysis on wind-wave data. The results show the tractability of this data to parametric methods, and the results are in agreement with previous spectral analyses of the data in the regions of overlap.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
air-sea interaction						
autoregressive integrated moving average model						
boundary layer turbulence						
joint probability density function analysis						
spectral analysis						
spectral filters						
stochastic model fitting						
wind-wave coupling						